

What is claimed is:

1. A method to reduce latency in accessing a memory from a bus, the method comprising:

3 pre-fetching a plurality of data from the memory to a cache queue in
4 response to a request; and
5 delivering the pre-fetched data from the cache queue to the bus
6 independently of the memory.

2. The method of claim 1 wherein pre-fetching comprises:
2 determining if an amount of data in the cache queue is above a predetermined
3 level; and
4 placing the request to a memory controller controlling the memory if the
5 amount of data is not above the predetermined level, the request causing the
6 memory controller to transfer the plurality of data to the cache queue, the
7 request being buffered in a request queue.

3. The method of claim 2 wherein delivering comprises:
2 transferring the data from the cache queue to the bus if the data in the
3 cache queue is ready.

1 4. The method of claim 1 further comprising:
2 determining if the request is valid; and
3 processing a cache miss request if the request results in a cache miss.

1 5. The method of claim 4 wherein processing the cache miss request
2 comprises:
3 providing a purge signal;
4 marking an entry in a scheduler according to the purge signal;
5 purging data corresponding to the marked entry; and
6 placing the request to the memory controller.

1 6. The method of claim 5 wherein the bus is a peripheral component
2 interconnect (PCI) bus.

1 7. The method of claim 6 wherein the request is one of a 32-byte and a
2 64-byte requests.

Sub
PC2
1

8. An apparatus to reduce latency in accessing a memory from a bus,
the apparatus comprising:

a pre-fetcher to pre-fetch a plurality of data from the memory to a cache
queue in response to a request; and

a cache controller coupled to the cache queue and the pre-fetcher to
deliver the pre-fetched data from the cache queue to the bus independently of
the memory.

9. The apparatus of claim 8 wherein the pre-fetcher comprises:

a watermark monitor to determine if an amount of data in the cache queue
is above a predetermined level;

a request packet generator coupled to the watermark monitor to place the
request to a memory controller controlling the memory if the amount of data is
not above the predetermined level, the request causing the memory controller to
transfer the plurality of data to the cache queue; and

a request queue coupled to the request packet generator to store the
request provided by the request packet generator.

Sub
A2

10. The apparatus of claim 9 wherein the cache controller transfers the
2 data from the cache queue to the bus if the data in the cache queue is ready.

11. The apparatus of claim 8 further comprising:

2 a peripheral bus controller coupled to the bus and the pre-fetcher to
3 determine if the request is valid;

4 a data coherence controller coupled to the pre-fetcher to provide a purge
5 signal when the request corresponds to a cache miss; and

6 a scheduler coupled to the request queue and the data coherence
7 controller to store entries corresponding to the requests, the entries being
8 marked according to the purge signal from the data coherence controller.

12. The apparatus of claim 11 further comprising:

2 a data mover coupled to the cache queue and the scheduler to transfer
3 data from the memory to the cache queue, the data mover purging data
4 corresponding to a marked entry from the scheduler.

13. The apparatus of claim 12 wherein the bus is a peripheral
2 component interconnect (PCI) bus.

Sub
A27

14. The apparatus of claim 13 wherein the request is one of a 32-byte
and a 64-byte requests.

15. A system comprising:

a memory;

a bus; and

a bus access circuit coupled to the memory and the bus to reduce latency
in accessing the memory from the bus, the circuit comprising:

a pre-fetcher to pre-fetch a plurality of data from the memory to a cache
queue in response to a request, and

a cache controller coupled to the cache queue and the pre-fetcher to
deliver the pre-fetched data from the cache queue to the bus independently of
the memory.

16. The system of claim 15 wherein the pre-fetcher comprises:

a watermark monitor to determine if an amount of data in the cache queue
is above a predetermined level;

a request packet generator coupled to the watermark monitor to place the
request to a memory controller controlling the memory if the amount of data is

6 not above the predetermined level, the request causing the memory controller to
7 transfer the plurality of data to the cache queue; and

8 a request queue coupled to the request packet generator to store the
9 request provided by the request packet generator.

1 17. The system of claim 16 wherein the cache controller transfers the
2 data from the cache queue to the bus if the data in the cache queue is ready.

1 18. The system of claim 15 wherein the bus access circuit further
2 comprises:

3 a peripheral bus controller coupled to the bus and the pre-fetcher to
4 determine if the request is valid;

5 a data coherence controller coupled to the pre-fetcher to provide a purge
6 signal when the request corresponds to a cache miss; and

7 a scheduler coupled to the request queue and the data coherence
8 controller to store entries corresponding to the requests, the entries being
9 marked according to the purge signal from the data coherence controller.

1 19. The system of claim 18 wherein the bus access circuit further
2 comprising:

Sub
Q2

3 a data mover coupled to the cache queue and the scheduler to transfer
4 data from the memory to the cache queue, the data mover purging data
5 corresponding to a marked entry from the scheduler.

1 20. The system of claim 19 wherein the bus is a peripheral component
2 interconnect (PCI) bus.

1 21. The system of claim 20 wherein the request is one of a 32-byte and a
2 64-byte requests.